BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PARALLEL PROCESSING SYSTEM AND DATA TRANSFERRING METHOD

Patent Number: JP5173991
Publication date: 1993-07-13
Inventor(s): OKABAYASHI ICHIRO
Applicant(s): MATSUSHITA ELECTRIC IND CO LTD
Requested Patent: ☐ JP5173991
Application Number: JP19920044399 19920302
Priority Number(s):
IPC Classification: G06F15/16; G06F13/38
EC Classification:
Equivalents: JP3389610B2

Abstract

PURPOSE:To enable a PE-to-PE communication and improve processor performance.
CONSTITUTION:The parallel processing system consists of plural PEs 1a, 1c, and 1d and a network 2 which connects the PEs mutually. The PE 1a is constituted by connecting a processor 3a, a memory 4a, and a data transfer device 5a to a common bus. The data transfer device 5 has three buffers and a data repeating device 6 has two buffers. Data from the PE1a to the PE1d are transferred in the order of the memory 4a, buffer 7a, buffer 10a, buffer 8c, buffer 11e, buffer 9d, and memory 4d as shown by a dotted line. Namely, the PE1c repeats the data by utilizing the buffer 8c. Consequently, neither memory writing nor reading is performed at the repeating PE at the time of an optional PE-to- PE communication, so the overhead at the repeating PE is reduced to improve the transfer performance. Further, the data transfer device performs no bus access, so the bus width is widened and the performance of the processor is improved.

AN

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号

特開平5－173991

(43)公開日　平成５年(1993)７月13日

(51)Int.Cl.⁵		識別記号	庁内整理番号	FI	技術表示箇所
G06F	15/16	400 N	9190－5L		
		320 V	8840－5L		
	13/38	340 C	9072－5B		

審査請求　未請求　請求項の数10(全 18 頁)

(21)出願番号　特願平4－44399

(22)出願日　平成４年(1992)３月２日

(31)優先権主張番号　特願平3－54529

(32)優先日　平３(1991)３月19日

(33)優先権主張国　日本（ＪＰ）

(71)出願人　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地

(72)発明者　岡林　一郎
大阪府門真市大字門真1006番地　松下電器産業株式会社内

(74)代理人　弁理士　小鍜治　明　（外２名）

(54)【発明の名称】　並列処理システムとデータ転送方法

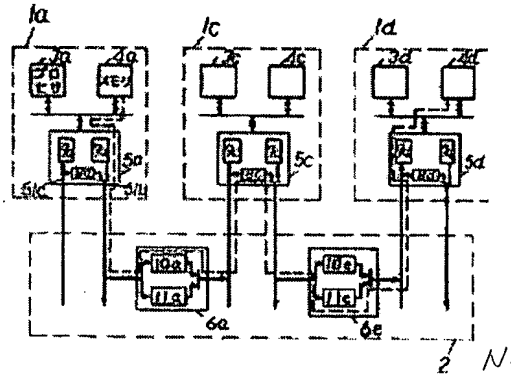
(57)【要約】

【目的】　本発明は並列処理システムに関するものであり、ＰＥ間通信及びプロセサ性能の向上を実現する。

【構成】　並列処理システムは複数のＰＥ1a・1c・1dと、ＰＥ間を相互に接続するネットワーク2で構成される。ＰＥ1aはプロセサ3a、メモリ4a、データ転送装置5aを共通のバスに接続した構成である。データ転送装置5は、３つのバッファを、データ中継装置6は２つのバッファを有する。ＰＥ1aからＰＥ1dへは、点線で示した様に、メモリ4a、バッファ7a、バッファ10a、バッファ8c、バッファ11c、バッファ9d、メモリ4dとデータを転送する。即ち、ＰＥ1cでバッファ8cを利用してデータを中継する。

【効果】　任意ＰＥ間通信時、中継ＰＥでのメモリライト／リードがないので、ここでのオーバーヘッドが軽減され、転送性能が向上する。また、ここでデータ転送装置がバスアクセスをしないので、バス幅も広がりプロセサの性能も向上する。

1　ＰＥ　Process Element
2　ネットワーク
3　プロセサ
4　メモリ
5　データ転送装置　data 전송장치
6　データ中継装置　〃 중계장치
7～11　バッファ



【特許請求の範囲】

【請求項 1】プロセサと、メモリと、第１，第２，第３の３つのバッファを有するデータ転送装置を具備する複数のプロセサエレメントと、前記複数のプロセサエレメント間で直接または１つ以上のプロセサエレメントを中継することで間接的にデータ転送が可能なように接続するネットワークを備え、データ転送の際に、送り手となるプロセサエレメントでは、データ転送装置がメモリまたはプロセサからデータを前記第１のバッファに取り込んでから前記ネットワークへ送出し、受け手となるプロセサエレメントでは、データ転送装置が前記ネットワークから前記データを前記第２のバッファに取り込んだ後、メモリまたはプロセサへ格納し、中継するプロセサエレメントでは、データ転送装置が前記ネットワークから前記データを前記第３のバッファに取り込んだ後、再び前記ネットワークへ送出することを特徴とする並列処理システム。

【請求項 2】第１，第２，第３の３つの入出力ポートと、前記第１の入出力ポートと第２の入出力ポートの間に接続された第１のバッファと、前記第１の入出力ポートと第３の入出力ポートの間に接続された第２のバッファと、前記第２の入出力ポートと前記第３の入出力ポートの間に接続された第３のバッファと、前記第１の入出力ポートから取り込んだデータにタグを付加するタグ生成部と、外部から前記第１の入出力ポートへデータを入力する際のアドレスを生成し、かつ入力回数をカウントするメモリアドレス生成部と、前記第１の入出力ポートから外部へデータを出力する際にデータ出力回数をカウントするカウンタと、前記第１の入出力ポートから出力するデータの一部をアドレスとして取り出し、データ出力の際にはこれを、データ入力の際には前記アドレス生成部のアドレスを選択して外部に出力する第１のセレクタと、前記第２の入出力ポートから外部へデータを出力する際に、外部アドレスを生成する第１の中継アドレス生成部と、前記第３の入出力ポートへ外部からデータを入力する際に、外部アドレスを生成する第２の中継アドレス生成部と、第２の入出力ポートから外部へデータを出力する際、前記第１のバッファと前記第３のバッファ出力の一部を選択する第２のセレクタと、前記第１のバッファからデータを出力する場合には前記第１の中継アドレス生成部の出力を、前記第３のバッファからデータを出力する場合には前記第３のバッファ出力の他の一部を選択する第３のセレクタとを具備することを特徴とするデータ転送装置。

【請求項 3】第１，第２，第３の３つの入出力ポートと、前記第１の入出力ポートと第２の入出力ポートの間に接続された第１のバッファと、前記第１の入出力ポートと第３の入出力ポートの間に接続された第２のバッファと、前記第２の入出力ポートと前記第３の入出力ポートの間に接続された第３のバッファと、前記第１の入出力ポートから取り込んだデータにタグを付加するタグ生成部と、外部から前記第１の入出力ポートへデータを入力する際のアドレスを生成し、かつ入力回数をカウントするメモリアドレス生成部と、前記第１の入出力ポートから外部へデータを出力する際にデータ出力回数をカウントするカウンタと、前記第１の入出力ポートから出力するデータの一部をアドレスとして取り出し、データ出力の際にはこれを、データ入力の際には前記アドレス生成部のアドレスを選択して外部に出力する第１のセレクタと、前記第２の入出力ポートから外部へデータを出力する際に、外部アドレスを生成する第１の中継アドレス生成部と、前記第３の入出力ポートから外部へデータを出力する際に、外部アドレスを生成する第２の中継アドレス生成部と、第２の入出力ポートから外部へデータを出力する際、前記第１のバッファと前記第３のバッファ出力の一部を選択する第２のセレクタと、前記第１のバッファからデータを出力する場合には前記第１の中継アドレス生成部の出力を、前記第３のバッファからデータを出力する場合には前記第３のバッファ出力の他の一部を選択する第３のセレクタと、第３の入出力ポートから外部へデータを出力する際、前記第２のバッファと前記第３のバッファ出力の一部を選択する第４のセレクタと、前記第２のバッファからデータを出力する場合には前記第２の中継アドレス生成部の出力を、前記第３のバッファからデータを出力する場合には前記第３のバッファ出力の他の一部を選択する第５のセレクタを具備することを特徴とするデータ転送装置。

【請求項 4】第１，第２，第３の３つの入出力ポートと、第１の入出力ポートと第２の入出力ポートの間に接続された第１のバッファと、前記第１の入出力ポートと第３の入出力ポートの間に接続された第２のバッファと、前記第２の入出力ポートと前記第３の入出力ポートの間に接続された第３のバッファと、前記第１の入出力ポートから取り込んだデータにタグを付加するタグ生成部と、外部から前記第１の入出力ポートへデータを入力する際のアドレスを生成し、かつ入力回数をカウントするメモリアドレス生成部と、前記第１の入出力ポートから外部へデータを出力する際にデータ出力回数をカウントするカウンタと、前記第１の入出力ポートから出力するデータの一部をアドレスとして取り出し、データ出力の際にはこれを、データ入力の際には前記アドレス生成部のアドレスを選択して外部に出力する第１のセレクタと、前記第２の入出力ポートから外部へデータを出力する際に、外部アドレスを生成する第１の中継アドレス生成部と、前記第３の入出力ポートから外部へデータを出力する際に、外部アドレスを生成する第２の中継アドレス生成部と、第２の入出力ポートから外部へデータを出力する際、前記第１のバッファと前記第３のバッファ出力の一部を選択する第２のセレクタと、前記第１のバッファからデータを出力する場合には前記第１の中継アドレス生成部の出力を、前記第３のバッファからデータを出力する場合には前記第３のバッファ出力の他の一部を選択する第３のセレクタと、第３の入出力ポートから外部へデータを出力する際、前記第２のバッファと前記第３のバッファ出力の一部を選択する第４のセレクタと、前記第２のバッファからデータを出力する場合には前記第２の中継アドレス生成部の出力を、前記第３のバッファからデータを出力する場合には前記第３のバッファ出力の他の一部を選択する第５のセレクタを具備することを特徴とするデータ転送装置。

【請求項 5】少なくとも３つの入出力ポートと、第１の入出力ポートと第２の入出力ポートの間に接続された第

1のバッファと、前記第1の入出力ポートと第3の入出力ポートの間に接続された第2のバッファと、前記第2の入出力ポートと前記第3の入出力ポートの間に接続された第3のバッファと、前記第1の入出力ポートから取り込んだデータにタグを付加するタグ生成部と、外部から前記第1の入出力ポートへデータを入力する際のアドレスを生成し、かつ入力回数をカウントするメモリアドレス生成部と、前記第1の入出力ポートから外部へデータを出力する際にデータ出力回数をカウントするカウンタと、前記第1の入出力ポートから出力するデータの一部をアドレスとして取り出し、データ出力の際にはこれを、データ入力の際には前記アドレス生成部のアドレスを選択して外部に出力する第1のセレクタと、第2の入出力ポートから外部へデータを出力する場合、前記第1のバッファ出力の一部と前記第3のバッファ出力の一部を選択する第2のセレクタと、前記第1のバッファ出力の他の一部と前記第3のバッファ出力の他の一部を選択する第3のセレクタと、第3の入出力ポートから外部へデータを出力する場合、前記第2のバッファ出力の一部と前記第3のバッファ出力の一部を選択する第4のセレクタと、前記第2のバッファ出力の他の一部と前記第3のバッファ出力の他の一部を選択する第5のセレクタとを具備することを特徴とするデータ転送装置。

【請求項 6】請求項 2又は4記載のタグ生成部は、データの種別を判定する制御情報部と、2回目以降に中継する中継部のアドレスを中継順に示す複数の中継アドレス部と、最終的なデータの格納アドレスを示すメモリアドレス部を有する情報をタグとして付加することを特徴とするデータ転送装置。

【請求項 7】請求項 3又は5記載のタグ生成部は、データの種別を判定する制御情報部と、中継する中継部のアドレスを中継順に示す複数の中継アドレス部と、最終的なデータの格納アドレスを示すメモリアドレス部を有する情報をタグとして付加することを特徴とするデータ転送装置。

【請求項 8】第1,第2の2つの入出力ポートと、Nを2以上の整数としてN個のバッファと、N入力1出力の出力セレクタと、N－1個の2入力1出力の入力セレクタを有し、前記出力セレクタの入力に前記N個のバッファ出力を接続し、出力を前記第1の入出力ポートに接続し、前記第2の入出力ポートを第1のバッファの入力及び前記N－1個の入力セレクタの入力の一端に接続し、第2から第Nまでのバッファ入力に前記入力セレクタの出力を接続し、Lを1以上N－1以下の整数として第Lのバッファ出力を第Lの入力セレクタの他端に接続することを特徴とするデータ中継装置。

【請求項 9】プロセサと、メモリと、請求項 2,3のいずれかに記載のデータ転送装置を共通のバスに接続した構成の複数のプロセサエレメントと、各格子点に請求項8記載のデータ中継装置を配してここを経由することである2つのプロセサエレメント間通信を行い、またNを2以上の整数として、N－1回プロセサエレメントを中継することで任意のプロセサエレメント間通信が可能となるネットワークを備え、データ転送の際に、送り手となるプロセサエレメントでは、データ転送装置がメモリまたはプロセサからデータを第1のバッファに取り込んでから前記ネットワークへ送出し、中継するプロセサエレメントでは、データ転送装置が前記ネットワークから前記データを第3のバッファに取り込んだ後、再び前記ネットワークへ送出し、受け手となるプロセサエレメントでは、データ転送装置が前記ネットワークから前記データを第2のバッファに取り込んだ後、メモリまたはプロセサへ格納し、また、Lを1以上N－1の整数として、第L回目のプロセサエレメントから第L＋1回目のプロセサエレメントへのデータ転送時に、前記データ中継装置の第Lのバッファを経由することを特徴とする並列処理システム。

【請求項 10】Nを2以上の整数として、少なくとも2つのポートを有するN個のプロセサエレメントと、K,Lを1以上N以下の整数として、N×N個の格子点を有し、各格子点を（K,L）とし、この格子点に少なくとも2つのポートを有するバッファ（K,L）を配したネットワークを備え、前記第Kのプロセサエレメントの一端をバッファ（K,L）の一端に共通に接続し、また前記バッファ（K,L）の他端をLが共通になるように接続し、この共通接続線を前記プロセサエレメントに接続するか、または外部ポートとする構成を基本単位として含む並列処理システムにおいて、前記第Kのプロセサエレメントからデータを送出する際に、バッファ（K,K）から順次データを送出するデータ転送方法。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、計算機分野でその将来性が期待されている並列処理システムに係わり、特にプロセッサエレメント間通信に関する。

【０００２】

【従来の技術】一般に並列処理システムは、計算処理を行なうプロセサエレメント（以下ＰＥ）と、ＰＥ間のデータ転送を行なうネットワークで構成される。

【０００３】以下図面を参照しながら、従来の並列処理システムにおけるＰＥ間通信の一例について説明する。図１０は従来の第１の並列処理システムの構成図、図１１は従来のデータ転送装置の構成図、図１２は従来のデータ中継装置の構成図を示すものである。これらは、電子情報通信学会・集積回路研究会ICD89-152に詳しく述べられている。ここでは、ＰＥ、ネットワークの一部を示す。また、データの流れを単方向に限って説明する。

【０００４】まず、従来の並列処理システム（図１０）について説明する。基本的には、ＰＥ1a, 1b, 1dとＰＥ間を相互に接続するネットワーク2で構成される。ＰＥ１

は全て同一の構成であり、ＰＥ1aを例にとれば、プロセサ3a，メモリ4a，データ転送装置5aを共通のバスに接続した構成である。また、データ転送装置5aは２つのバッファ7a，9aを有する。また、ネットワーク2内部にデータ中継装置6a，6eを設ける。データ中継装置6a・6eはそれぞれバッファ10a，10eを有する。ネットワーク2は、任意ＰＥ間通信が第３のＰＥを１回経由することで可能なもの（ＰＥ間距離が２）である。以上の従来の並列処理システム において、ＰＥ1aからＰＥ1dへのデータの流れは、メモリ4a、バッファ7a、バッファ10a、バッファ9c、メモリ4c 、バッファ7c、バッファ10e、バッファ9d、メモリ4d となる。これを図１０中に点線で示す。

【０００５】続いて、従来のデータ転送装置（図１１）について説明する。入出力ポート17aはメモリ4に、入出力ポート17b，17cはネットワーク2に接続される。入出力ポート17aから17bへのデータの流れは次の様になる。メモリアドレス生成部12aよりセレクタ18a経由で、アドレス50aを出力してメモリリードを行ないデータ51aを入出力ポート17aからバッファ7 へ取り込む。次に、中継アドレス生成部15aからアドレス50bを出力して、入出力ポート17bからデータ51bを出力する。

【０００６】また、入出力ポート17cから17aへのデータの流れは次の様になる。中継アドレス生成部15bからアドレス50cを出力して、入出力ポート17cからデータ51cを入力し、バッファ9に取り込む。次に、入出力ポート17aから、メモリアドレス生成部12bよりセレクタ18a経由でアドレス50aを、バッファ9よりデータ51a を出力してメモリライトを行なう。なお、制御部16a・16bはバッファ状態52a・52bを監視する。

【０００７】さらに続いて、従来のデータ中継装置（図１２）について説明する。データ51bはバッファ10に格納される。制御部31aはバッファ10のリード／ライトを制御する。デコーダ30a・30bはアドレス50b・50cを監視し、自分がアクセスされた際に、トライステートバッファ32a・32cをイネーブルとして、バッファ状態52a・52bを外部へ通過させる。ここでのバッファ状態とは、書き込み側はバッファフル、読みだし側はバッファエンプティに関するものである。

【０００８】次に、図１３は従来のデータ転送方法を示す図である。これは、ネットワーク2が完全クロスバ網の例である。ＰＥからのデータ送出順序をデータ中継装置6a-6p内に示す。即ち、最初のステップでＰＥ1aはデータ中継装置6aに、1bは6e、1cは6i、1dは6mに一斉にデータを送出する。次のステップでは、ＰＥ1aはデータ中継装置6bに、1bは6f、1cは6j、1dは6nに一斉にデータを送出する。以下同様で、端まで送出し終わると最初に戻る。最初のステップ終了後、ＰＥ1aがデータ中継装置6aよりデータを受信する。

【０００９】最後に、図１４は従来の第２の並列処理システム の構成図である。これは、電子情報通信学会・コンピュータシステム 研究会CPSY89-1に詳しく述べられている。ここで、ＰＵ（プロセッシングユニット）は図１４（a）に示す様にメッシュ状に接続されている。各ＰＵは図１４（b）に示す様にＣＰＵ71、ローカルメモリ72、周辺ＬＳＩ73を共通のバスに接続した構成である。また、４つのポート75a-dを有し、２ポートＲＡＭであるコネクションメモリ74a-bを介して他ＰＵと通信を行なう。

【００１０】

【発明が解決しようとする課題】しかしながら上記の様な第１の並列処理システム では、中継するＰＥがデータを一旦メモリにライトしてから再度リードするので、ここでのオーバーヘッドが大きいという問題点を有していた。またメモリアクセスを行なうので、バスネックが発生しプロセサ性能も低下する。

【００１１】また、上記の様なデータ転送方式では、最初のステップ終了後ＰＥ1aのみがデータ中継装置6a，6e，6i，6mに接続されているため受信可能で、このバスだけに負荷が集中し、システム 全体の転送性能は低下する。

【００１２】また、上記の様な第２の並列処理システムでは、全ＰＵが同期して隣接するＰＵと通信する場合は非常に高速であるが、距離の遠いＰＵとの通信は遅い。任意のＰＥ間距離はＮ×Ｎのシステム で最大Ｎ、平均Ｎ／２である。また個々のＰＵで通信要求がランダム に発生する場合の対応、あるいは他ネットワークへの拡張の２点でも不利である。

【００１３】本発明は上記問題点に鑑み、高いプロセサ性能、高速なＰＥ間通信を実現し、かつ柔軟な並列処理システム を提供することを目的とする。

【００１４】

【課題を解決するための手段】上記問題点を解決するために、本発明の並列処理システム は、プロセサと、メモリと、第１，第２，第３の３つのバッファを有するデータ転送装置を具備する複数のプロセサエレメントと、前記複数のプロセサエレメント間で直接または１つ以上のプロセサエレメントを中継することで間接的にデータ転送が可能なように接続するネットワークを備えたものであり、データ転送の際に、送り手となるプロセサエレメントでは、データ転送装置がメモリまたはプロセサからデータを第１のバッファに取り込んでから前記ネットワークへ送出し、受け手となるプロセサエレメントでは、データ転送装置が前記ネットワークから前記データを第２のバッファに取り込んだ後、メモリまたはプロセサへ格納し、中継するプロセエレメントでは、データ転送装置が前記ネットワークから前記データを第３のバッファに取り込んだ後、再び前記ネットワークへ送出するものである。

【００１５】また、本発明のデータ転送方式は、少なくとも２つのポートを有するＮ個（Ｎは２以上の整数）のプロセサエレメントと、Ｎ×Ｎ個の格子点を有し、各格

子点を（K，L）（K，Lは１以上N以下の整数）とし、ここに少なくとも２つのポートを有するバッファ（K，L）を配したネットワークを備え、前記第Kのプロセサエレメントの一端をバッファ（K，L）の一端に共通に接続し、またバッファ（K，L）の他端をLが共通になるように接続し、この共通接続線を前記プロセサエレメントに接続するか、または外部ポートとする構成を基本単位として含む並列処理システム において、前記第Kのプロセサエレメントからデータを送出する際に、バッファ（K，K）から順次データを送出するものである。

【００１６】

【作用】本発明の並列処理システム では、上記した構成によって、ＰＥ間で通信を行なう場合に、送り手となるプロセサエレメントでは、データ転送装置がメモリまたはプロセサからデータを第１のバッファに取り込んでからネットワークへ送出する。また、受け手となるプロセサエレメントでは、データ転送装置がネットワークからデータを第２のバッファに取り込んだ後、メモリまたはプロセサへ格納する。さらに、中継するプロセエレメントでは、データ転送装置がネットワークからデータを第３のバッファに取り込んだ後、再びネットワークへ送出する。従って、中継ＰＥでのメモリライト／リードがないので、ここでのオーバーヘッドが軽減され、転送性能が向上する。また、データ転送装置がバスアクセスをしないので、バス幅も広がりプロセサの性能も向上する。

【００１７】また、本発明のデータ転送方式では、Ｎ×Ｎ個（Ｎは２以上の整数）の格子点を有するクロスバ網に対して、第Ｋのプロセサエレメントは、格子点（Ｋ，Ｋ）から順次データを送出するので、受信の際特定のパスにトラフィックが集中することはなく、転送性能が向上する。

【００１８】

【実施例】以下本発明の実施例について、図面を参照しながら説明する。

【００１９】これらの実施例では説明及び図面の簡単化のため、データの流れを一方向に限定する。また、ネットワークは、任意ＰＥ間通信が第３のＰＥを１回経由することで可能なもの（ＰＥ間距離が２）である。

【００２０】本発明の第１の実施例における並列処理システム の全体構成図（図６）を説明する。これはＰＥ数４の並列処理システム である。図６の構成によりＰＥ間距離２を実現する。図１はこの抜粋と考えてよく、ＰＥ1a、データ中継装置6a、ＰＥ1c、データ中継装置6e、ＰＥ1dと接続されることがわかる。

【００２１】図１は本発明の第１の実施例における並列処理システム の構成図である。基本的には、ＰＥ1a・1c・1dとＰＥ間を相互に接続するネットワーク2で構成される。制御線を含めた全体的な接続は、図７を用いて後ほど説明する。

【００２２】ＰＥは全て同一の構成であり、ＰＥ1aを例にとれば、プロセサ3a、メモリ4a、データ転送装置5aを共通のバスに接続した構成である。また、データ転送装置5aは３つのバッファ7a・8a・9aを有する。また、ネットワーク2内部にデータ中継装置6a・6eを設ける。データ中継装置6aはバッファ10a・11aを、6eはバッファ10e・11eを有する。以上の並列処理システム において、ＰＥ1aからＰＥ1dへのデータの流れは、メモリ4a、バッファ7a、バッファ10a、バッファ8c、バッファ11e、バッファ9d、メモリ4dとなる。これを図１中に点線で示す。

【００２３】ここで、データを中継するＰＥ1cにおいて、データ転送装置5cはデータ中継装置6aから受信したデータを内部のバッファ8c経由でデータ中継装置6eに送出する。ここで、メモリ4cのアクセスを伴わないので、転送速度の向上及びバス幅の拡大の両者が同時に可能となる。

【００２４】以下、本実施例を実現するための要素技術であるデータ転送装置、データ中継装置等について順次説明する。

【００２５】まず、第１のデータ転送装置について説明する。図２は、本発明の第１の実施例におけるデータ転送装置の構成図である。入出力ポート17aはメモリ4に、17b，17cはネットワーク2のバッファ中継装置6に接続される。

【００２６】入出力ポート17aから17bへのデータの流れは次の様になる。メモリアドレス生成部12aよりセレクタ18a経由で、アドレス50を出力してメモリリードを行なう。データ51を入出力ポート17aから取り込み、タグ生成部13の出力をタグとしてデータに付加した後、バッファ7へ取り込む。次に、中継アドレス生成部15aからセレクタ18c経由でアドレス50b、バッファ部7からセレクタ18b経由でデータ51bを入出力ポート17bより出力する。ここで、メモリアドレス生成部12aはリード回数のカウントも行なう。タグについては、図４を用いて後ほど説明する。

【００２７】また、入出力ポート17cから17aへのデータの流れは次の様になる。中継アドレス生成部15bからアドレス50cを出力して、入出力ポート17cからデータ51cを入力し、バッファ9に取り込む。次に、入出力ポート17aから、バッファ9 の出力の一部をセレクタ18a経由でアドレス50に、バッファ9の他の一部をデータ51に出力してメモリライトを行なう。ここで、カウンタ14はライト回数のカウントを行なう。

【００２８】最後に、入出力ポート17cから17bへのデータの流れは次の様になる。中継アドレス生成部15bからアドレス50cを出力して、入出力ポート17cからデータ51cを入力し、タグ変換部131でタグ部分を変換したのちバッファ8に取り込む。次に、アドレス50bとしてバッファ8の出力の一部をセレクタ18c経由で、データ51bとしてバッファ部8の出力の他の一部をセレクタ18b経由でそれぞれ入出力ポート17bより出力する。即ち中継アドレス

として、メモリからネットワークへの転送時は中継アドレス生成部15a出力を、ネットワークからネットワークへの転送時はデータの一部を用いる。またメモリアドレスとして、リード時はメモリアドレス生成部18a出力を、ライト時はデータの一部を用いる。

【００２９】なお、制御部16a、16bはデータ中継装置のバッファ状態52a、52bを監視する。ここでは、入出力ポート17b,17cを単方向としたが、これは双方向でもよい。図15はこれを示したもので、第１の実施例におけるデータ転送装置を双方向にした場合の構成図である。この場合はバッファ7、8、9、内部線を双方向化した上で、セレクタ18e,18dを入出力ポート17c側に設ける。タグ変換部131は１つであるので、データの流れが入出力ポート17c->入出力ポート17bの場合はバッファ8入力時、入出力ポート17b->入出力ポート17cの場合はバッファ8出力時にタグの変換をすることになる。

【００３０】次に、第２のデータ転送装置について説明する。図３は本発明の第２の実施例におけるデータ転送装置の構成図である。基本的には図２と同様であるが、入出力ポート17aから17bへのデータの流れに若干の相違点があるのでこれについて説明する。メモリアドレス生成部12aよりセレクタ18a経由で、アドレス50を出力してメモリリードを行なう。データ51を入出力ポート17aから取り込み、タグ及び中継アドレス生成部130の出力をタグとしてデータに付加した後、バッファ7へ取り込む。次に、アドレス50bとしてバッファ部7の出力の一部をセレクタ18c経由で、データ51bとしてバッファ部7の他の一部をセレクタ18b経由でそれぞれ入出力ポート17bより出力する。即ち中継アドレスとして、メモリからネットワークへの転送時、ネットワークからネットワークへの転送時共にデータの一部を用いる。図２ではデータ中継装置15aで生成したアドレスを、図３ではタグ及び中継アドレス生成部130で生成することになる。

【００３１】ここでは、入出力ポート17b,17cを単方向としたが、これは双方向でもよい。図16はこれを示したもので、第２の実施例におけるデータ転送装置を双方向にした場合の構成図である。この場合はバッファ7、8、9、内部線を双方向化した上で、セレクタ18e,18dを入出力ポート17c側に設ける。タグ変換部131は１つであるので、データの流れが入出力ポート17c->入出力ポート17bの場合はバッファ8入力時、入出力ポート17b->入出力ポート17cの場合はバッファ8出力時にタグの変換をすることになる。

【００３２】次に、タグ生成部13で生成するタグについて、本実施例におけるデータ形式の構成図である図４を用いて説明する。データはデータ部24とタグ部で構成される。タグは、データの種別（属性・形式等）を示す制御情報部20、現在の転送が第何ステップであるかを示す回数部21、２回目以降に中継する中継部のアドレスを中継順に示す複数の中継アドレス部22a,22b、最終的なデータの格納アドレスを示すメモリアドレス部23を備える。図４（ａ）が送出、中継ＰＥ間、図４（ｂ）が中継、受信ＰＥ間のデータの形式である。図１の例では図４（ａ）がＰＥ1a-1c間、図４（ｂ）がＰＥ1c-1d間のデータの形式である。

【００３３】図１を用いてタグとＰＥ間通信の関係について説明する。データ転送装置は図３の構成とする。データ転送装置5aのタグ及び中継アドレス生成部130で、制御情報部20に制御情報を、回数部21には回数を、中継アドレス部22aにはデータ中継装置6aのアドレスを、中継アドレス部22bにはデータ中継装置6eのアドレスを、メモリアドレス部23にはメモリ4dのアドレスをそれぞれ示したタグを生成・付加する。これが図４（ａ）である。例として、制御情報部20はデータ長を示す３、回数部21は一回目を示す１、中継アドレス部22aはデータ中継装置6aが手前にあるので０、中継アドレス部22bも０を付加する。データ転送装置5aは、中継アドレス部22a`０`を送出してデータ中継装置6aへデータを送出する。ここで制御情報部20の３はデータ長が２の３乗つまり８ワードである旨等を含む。次にデータ転送装置5cのタグ変換部131ではデータ中継装置6aからデータを取り込み、回数部21を２回目を示す２にし、中継アドレス部22aを削除して図４（ｂ）に示す形式に変換後、中継アドレス部22b`０`を送出してデータ中継装置6eへデータを送出する。最後に、データ転送装置5dではデータ中継装置6eからデータを取り込んだ後、メモリアドレス部23を出力してメモリ4dへデータをライトする。データ部が８ワードであるので、メモリアドレス部23に示すアドレスより順次８ワードライトすることになる。

【００３４】ここで、データ取り込みに関して、候補となるデータ中継装置が複数ある場合は、順次スキャンしてデータの準備されたものから取り込めばよい。これら一連の動作で、データ転送装置はネットワークから受けたデータを再度ネットワークに送出するか、メモリにライトするか決める必要がある。これは例えば、回数部21を見て判断すればよい。また、別の方法としてデータ中継装置にＰＥ間距離に相当する複数のバッファを準備し、ある特定のバッファからのデータはメモリ、その他はネットワークと決めれば回数部21は不要となる。これは、図５にて後ほど説明する。

【００３５】また、データ転送装置が図２の構成の場合には、図４の中継アドレス部22aは存在せず、代わりに中継アドレス生成部15aが一回目の中継アドレスを生成することになる。なお、タグの構成としては回数部21の代わりに、複数の中継アドレス部の最後に終了符号を付ける形式も可能である。

【００３６】さらに別の方法として、タグ変換をせず図４（ａ）の形式のままで、１回目は22aを、２回目は22bを中継アドレスとして送出する方法もある。この場合は図２等におけるタグ変換部131は不要であるが、セレク

タ18cで１、２回目で選択するビット位置を変える必要が生じる。また少しではあるが、タグのビット数が大きくなる。

【００３７】データ毎にアドレス情報が付加された以上の構成により、送り手及び受け手のＰＥが複数で、かつデータ数が複数で、また流れる順序がランダムな場合でも、複雑な制御なしで、確実に転送が実現できる。

【００３８】さて次に、データ中継装置について説明する。図５は本発明の実施例におけるデータ中継装置の構成図である。データ中継装置には２つのモードを有する。

【００３９】第１のモードは、１本のバッファとして動作するもので、入力セレクタ35はバッファ10出力を、出力セレクタ34はバッファ11出力を選択する。バッファ10及び11を連続した１本のバッファとして使用する。動作は従来例（図12）と同様である。即ち、制御部31aはバッファ10、11のリード／ライトを制御する。デコーダ30a・30bはアドレス50b・50cを監視し、自分がアクセスされた際に、トライステートバッファ32a・32bをイネーブルとして、バッファ状態52a・52bを外部へ通過させる。この時、制御部31b，トライステートバッファ32c・32dはディスエーブルである。

【００４０】第２のモードは、２本のバッファを並列に動作させるもので、入力セレクタ35は入出力ポート36aを、出力セレクタ34は必要に応じてバッファ10または11出力を選択する。バッファ10及び11を独立な２本のバッファとして使用する。

【００４１】制御部31aはバッファ10のリード／ライトを、制御部31bはバッファ11のリード／ライトを制御する。デコーダ30a・30bはアドレス50b・50cを監視し、自分がアクセスされた際に、トライステートバッファ32a-32dをイネーブルとして、バッファ状態52a-52dを外部へ通過させる。ここでのバッファ状態とは、書き込み側はバッファフル、読みだし側はバッファエンプティに関するものであり、バッファ状態52a・52bがバッファ10に、バッファ状態52c・52dがバッファ11にそれぞれ対応する。

【００４２】第２のモードでは、データ中継装置6にＰＥ間距離２に相当する２本のバッファが存在することになる。データ中継装置6a・6eでは、バッファ10a・10eが１回目、バッファ11a・11eが２回目のデータを格納する。従って、データは前記した様に図１の点線の流れとなる。

【００４３】データ転送装置5は、データ中継装置の２つのバッファ状態を監視して、送受可能な方とデータのやりとりを行なう。ここで、データ転送装置5cがバッファ10aから取り込んだ時はバッファ8cに、バッファ11aから取り込んだ時はバッファ9cにそれぞれ格納する制御を行なうことで、図４のタグの回数部21は不要となる。

【００４４】第３のＰＥを介するＰＥ間通信時、本データ中継装置の第１のモードを用いるデータ中継装置を用いた場合はデッドロックが発生するが，本データ中継装置の第２のモードを用いることで、１回目と２回目のデータが独立に扱えるので、デッドロックが回避できるが、ここの事情について説明する。

【００４５】図17は本発明のデータ中継装置（第１のモード）を用いた場合の転送の様子を示す図、図18は本発明のデータ中継装置（第２のモード）を用いた場合の転送の様子を示す図である。ここでは、ＰＥ1b，1a，1c間でデータが流れる場合について考える。また簡単化のためバッファ7，8，9は１段、図17でバッファ10d，10a，10fは２段、図18でバッファ10d，10a，10f，11d，11a，11fは１段とする。

【００４６】送信、中継、受信ＰＥとデータの関係は次の様になる。

送信ＰＥ1b->中継ＰＥ1a->受信ＰＥ1c　：　データc1，c2，c3，c4

送信ＰＥ1a->中継ＰＥ1c->受信ＰＥ1b　：　データb1，b2，b3，b4

送信ＰＥ1c->中継ＰＥ1b->受信ＰＥ1a　：　データa1，a2，a3，a4

データ中継装置（第１のモード）を用いた場合では、図17に示す様な状態に陥った場合にデッドロックとなる。ここで例えばＰＥ1bはデータc4またはa2を送出したいが、データ中継装置6dのバッファ10dがフルであるので送れない。データ中継装置6dはデータc3を吐き出したいがc3が入るべきバッファ8aがフルであるので転送できない。バッファ8aに空きが生じるためにはバッファ10aに空きが生じる必要があるが、データb3が入るべきバッファ8cがフルであるのでバッファ10aは空かない。バッファ8cが空くためにはバッファ10fが空く必要があるが、このためにはバッファ8bが空く必要がある。そのためにはバッファ10dが空く必要があり、結局どのバッファも空くことはない、つまりデッドロックとなる。この様に複数のＰＥで閉じたループを構成する場合にデッドロックが発生する可能性が高い。

【００４７】データ中継装置（第２のモード）を用いた場合では、図17に相当する状態が図18(a)である。例えばデータ中継装置6dについてみればバッファ10dに１回目の転送途中のデータc3、バッファ11dに２回目の転送途中のデータa1が格納される。

【００４８】次のサイクルではデータa1がバッファ9a、データc1がバッファ9c、データb1がバッファ9bに転送される。この状態を示したのが図18(b)である。バッファ9a，9c，9bのデータはメモリにライトされるのでこれらのバッファにはすぐ空きが生じる。こうなると例えばＰＥ1bはデータa2をバッファ11d経由でバッファ9aに送れる。バッファ8bに空きが生じるのでバッファ10fのデータa3がバッファ8bに転送可能となる。以下同様に順次データが流れデッドロックは生じない。

【００４９】以上により、ランダム な通信要求が発生した場合でも確実に動作できる。また１回目と２回目のデータの優先度付けを適切に行なうことで転送性能も向上する。また直接ＰＥ間で転送する場合は、第１のモードで大きなバッファリングが可能となる。

【００５０】データ転送装置とデータ中継装置間の制御線を含めた信号線の接続の様子を図７に示す。データ中継装置6aと6b、6aと6cの信号線が共通に接続される。中継アドレスによりアクティブなデータ中継装置が選択されてデータ・アドレスが受け渡される。またバッファ状態は選択されたデータ中継装置のみが出力し、他のデータ中継装置はハイインピーダンスである。

【００５１】次にネットワークでのアドレス・データの形式を、本発明の実施例におけるアドレス・データ構成図である図８を用いて説明する。ここでは、図８（ａ），（ｂ）２つの例を示す。一般に並列処理システム ではメモリとネットワークでのバス幅が異なり、ネットワーク側が狭くなる。そのため、ネットワークとのインターフェースでデータ幅の変換が必要となる。

【００５２】図８（ａ）では、データ転送装置1aからデータ中継装置6a間のデコーダ30aへアドレス50bが入力される。データ51b（アドレス以外）は、データ転送装置1aの出力ラッチ40に格納された後、セレクタ41で分解されて、データ中継装置6aへ入力される。図８（ｂ）では、アドレス50b及びデータ51bはデータ転送装置1aの出力ラッチ40に格納された後、セレクタ41で分解されて、データ中継装置6aへ入力され、アドレス50bはデコーダ30aへ、データ51bは内部へ入力される。図８（ｂ）は、データにアドレス情報が含まれ、データ中継装置は入力データを常に監視し、自分のアドレスに対応するものが出現した場合にデータを取り込むデーターフロー的な制御となる。図８（ａ）に比べて、制御ロジックは複雑になり、転送量も多いが、データ転送装置とデータ中継装置間の配線数は少なくなる。

【００５３】最後に、本発明の実施例におけるデータ転送方法について図９、図１９で説明する。図９は本発明の実施例におけるデータ転送方法を示す図、図１９は同実施例における時間と転送レートの関連図である。図９は、ネットワーク2が完全クロスバ網の例である。ＰＥからのデータ送出順序をデータ中継装置6a-6p内に示す。即ち、最初のステップでＰＥ1aはデータ中継装置6aに、1bは6f、1cは6k、1dは6pに一斉にデータを送出する。次のステップでは、ＰＥ1aはデータ中継装置6bに、1bは6g、1cは6l、1dは6mに一斉にデータを送出する。以下同様で、端まで送出し終わると最初に戻る。これにより最初のステップ終了後、全てのＰＥで受信が可能となる。即ち、ＰＥ1aはデータ中継装置6a、1bは6f、1cは6k、1dは6pよりそれぞれデータを受信できる。従って特定のネットワークの負荷が偏らないので、システム 全体の転送効率が向上する。これを図１９に示す。従来は、時間１で１つのチャネルでのみ送受が行なわれるので転送レートは１である。時間２で転送レート２、３で３、４で４と増えて行きその後減少し、時間７で終了する。これを点線で示す。本実施例では実線で示した様に時間１－４で全チャネルが動作つまり転送レートが４であり、時間４で転送は終了する。

【００５４】なお、本例は既に説明した図６の様に部分的にクロスバを有するシステム に適用可能である。また、最初に述べた様に、ここではデータの流れを一方向に限定したが、バッファの双方向化、セレクタなど一部の回路を２つ持つことで、二方向の流れにも容易に対応できる。

【００５５】また、ＰＥ間距離が２の並列処理システムについて述べたが、タグの中継アドレス、データ中継装置のバッファ本数をＮとすることで、ＰＥ間距離がNの並列処理システム に拡張可能である。またこれらを組み合わせて、各種の形態のネットワークを実現することが可能となる。

【００５６】

【発明の効果】以上述べてきた様に、本発明の並列処理システム では、任意ＰＥ間通信時、中継ＰＥでのメモリライト／リードがないので、ここでのオーバーヘッドが軽減され、転送性能が向上する。また、データ転送装置がバスアクセスをしないので、バス幅も広がりプロセサの性能も向上する。

【００５７】また本発明のデータ転送方式では、ネットワークの負荷が分散する、つまり受信側のＰＥが均等に動作できるので、システム 全体の転送性能が向上する。

【００５８】さらに、本発明のデータ転送装置、データ中継装置を用いることで、各種のネットワークを有する並列処理システム が構成できる。

【００５９】単体プロセサの計算機性能及び半導体技術の限界が見えてきた現在、並列処理システム への期待は非常に大きく、本発明は極めて有用なものである。

【図面の簡単な説明】

【図１】本発明の第１の実施例における並列処理システム の構成図

【図２】本発明の第１の実施例におけるデータ転送装置の構成図

【図３】本発明の第２の実施例におけるデータ転送装置の構成図

【図４】本発明の実施例におけるデータ形式の構成図

【図５】本発明の実施例におけるデータ中継装置の構成図

【図６】本発明の第１の実施例における並列処理システム の全体構成図

【図７】同実施例における接続詳細図

【図８】本発明の実施例におけるアドレス・データ構成図

【図９】本発明の実施例におけるデータ転送方法を示す

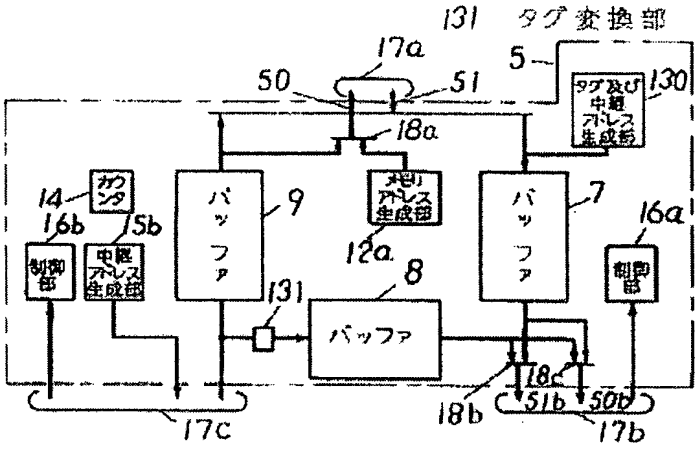
図

【図１０】従来の第１の並列処理システム の構成図
【図１１】従来のデータ転送装置の構成図
【図１２】従来のデータ中継装置の構成図を示す図
【図１３】従来のデータ転送方法を示す図
【図１４】従来の第２の並列処理システム の構成図
【図１５】第１の実施例におけるデータ転送装置を双方向にした場合の構成図
【図１６】第１の実施例におけるデータ転送装置を双方向にした場合の構成図
【図１７】データ中継装置（第１のモード）を用いた場合の転送の様子を示す図
【図１８】データ中継装置（第２のモード）を用いた場合の転送の様子を示す図
【図１９】本発明の実施例におけるデータ転送方法における時間と転送レートの関連図

【符号の説明】

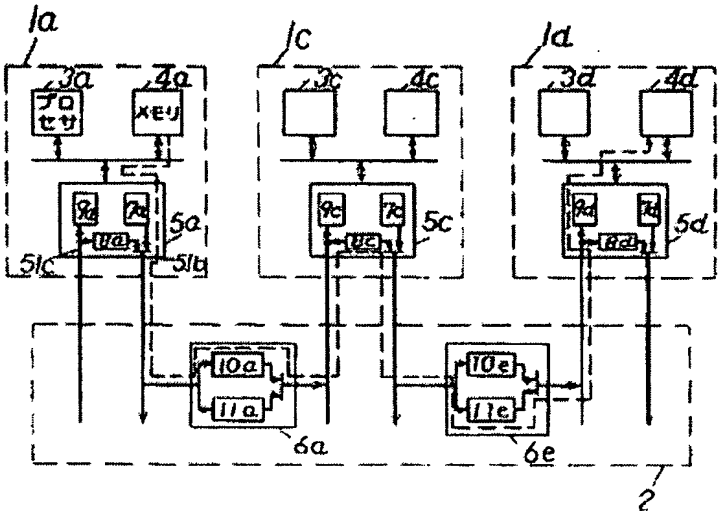
１　ＰＥ（プロセサエレメント）
２　ネットワーク
３　プロセサ
４　メモリ
５　データ転送装置
６　データ中継装置
７－１１　バッファ
１２　メモリアドレス生成部
１３　タグ生成部
１４　カウンタ
１５　中継アドレス生成部
１６　制御部
１７　入出力ポート
１８　セレクタ
２０　制御情報部
２１　回数部
２２　中継アドレス部
２３　メモリアドレス部
２４　データ部
３０　デコーダ
３１　バッファ制御部
３２　トライステートバッファ
３４　出力セレクタ
３５　入力セレクタ
３６　入出力ポート
４０　出力ラッチ
４１　セレクタ
５０　アドレス
５１　データ
５２　バッファ状態
７０　ＰＵ（プロセッシングユニット）
７１　ＣＰＵ
７２　ローカルメモリ
７３　周辺ＬＳＩ
７４　コネクションメモリ
７５　ポート
１３０　タグ及び中継アドレス生成部
１３１　タグ変換部

【図３】

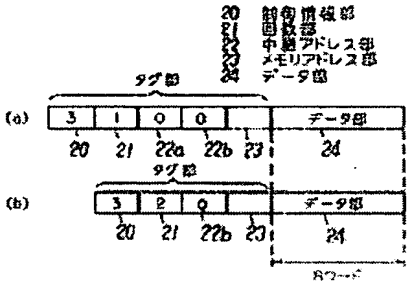


【図1】

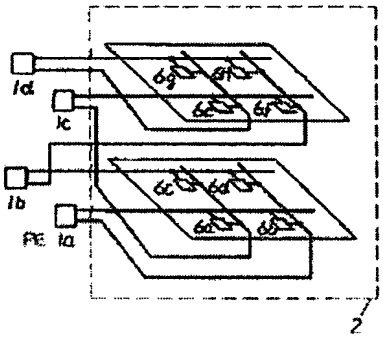
1	ＰＥ
2	ネットワーク
3	プロセサ
4	メモリ
5	データ転送装置
6	データ中継装置
7~11	バッファ



【図4】

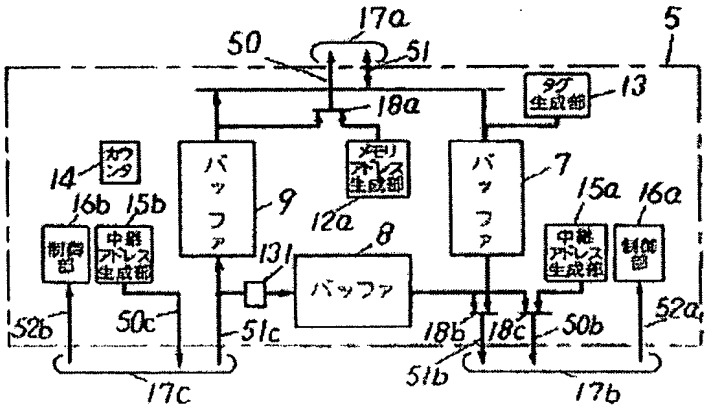


【図6】



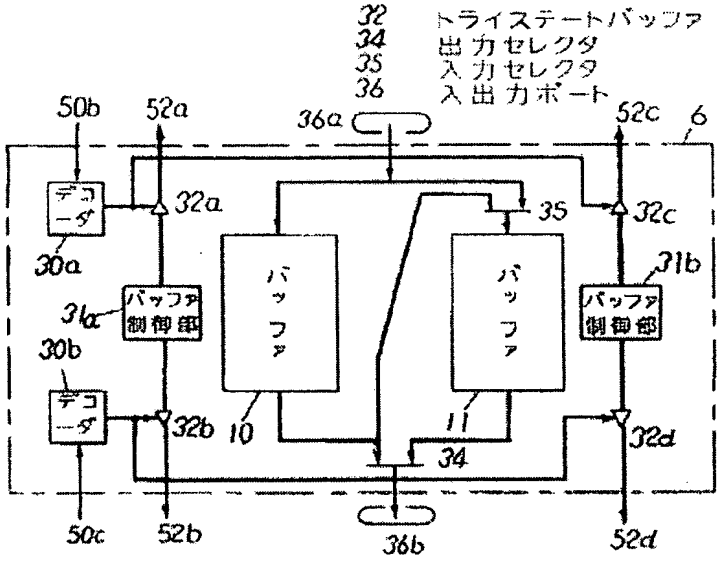
【図2】

17　入出力ポート
18　セレクタ
50　アドレス
51　データ
52　バッファ状態

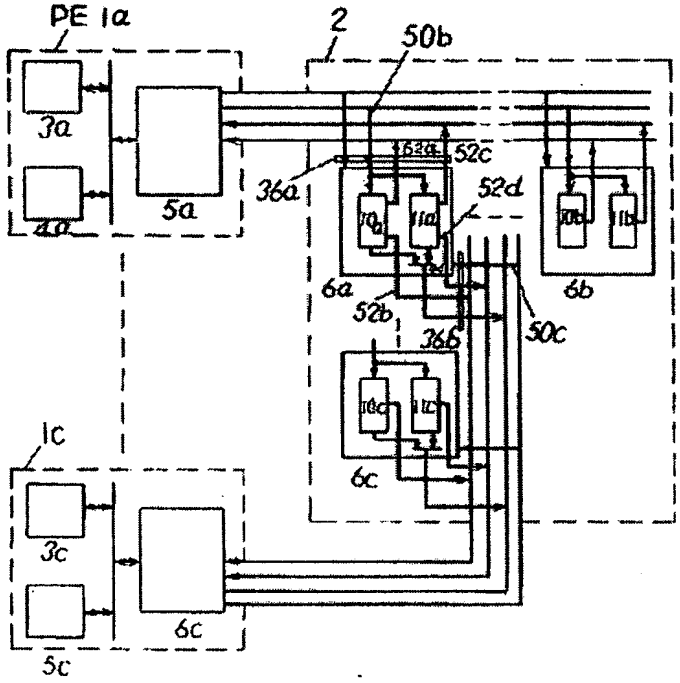


【図5】

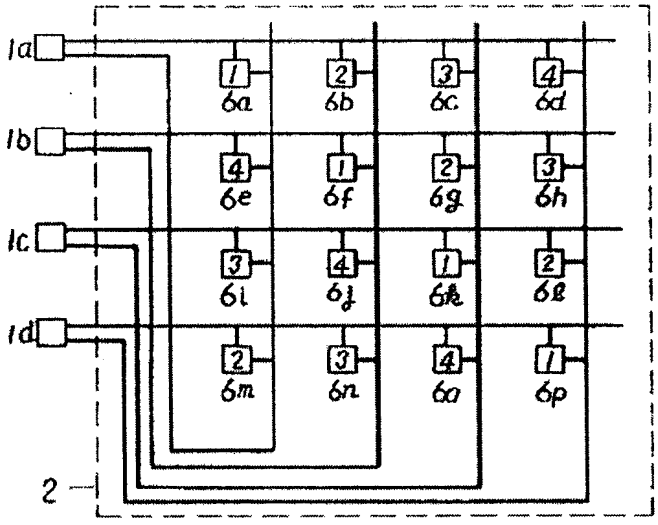
32　トライステートバッファ
34　出力セレクタ
35　入力セレクタ
36　入出力ポート



【図7】

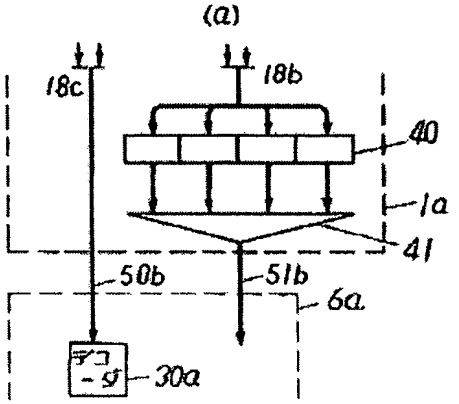

PE 1a
2
50b
3a
4a
5a
52c
36a
52d
6a
6b
52b
36b
50c
1c
3c
6c
5c

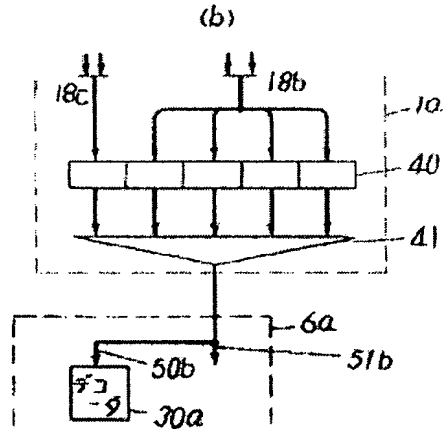
【図9】


1a
1b
1c
1d
2
6a
6b
6c
6d
6e
6f
6g
6h
6i
6j
6k
6l
6m
6n
6o
6p

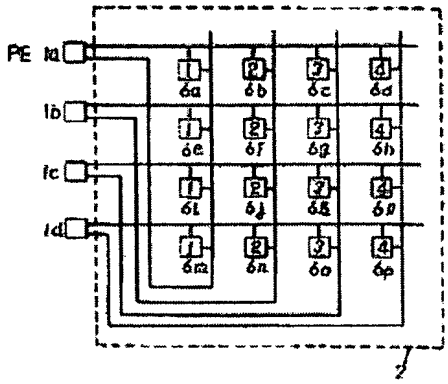
【図８】

40　出力ラッチ
41　セレクタ



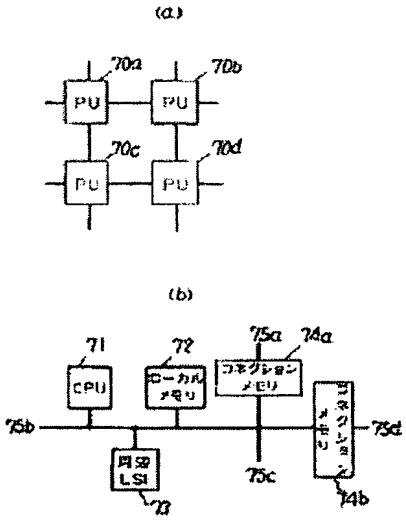


【図１３】

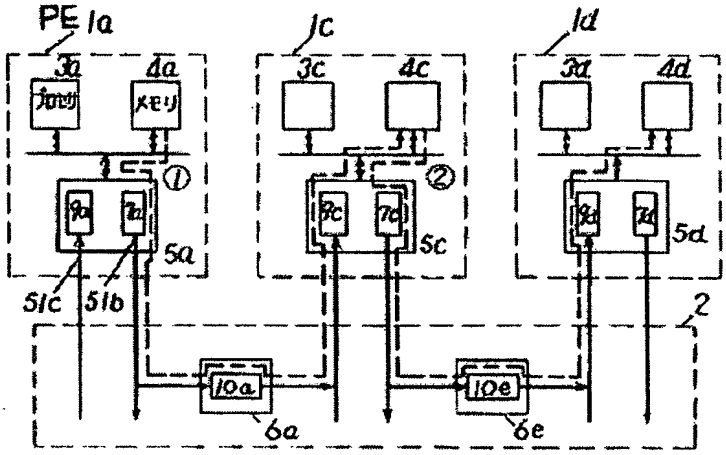


【図１４】

75　ポート

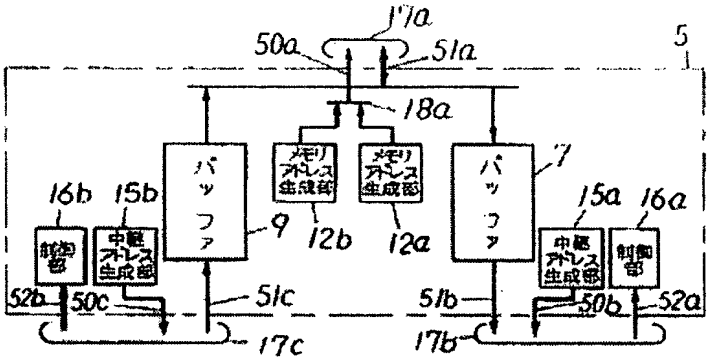


【図１０】

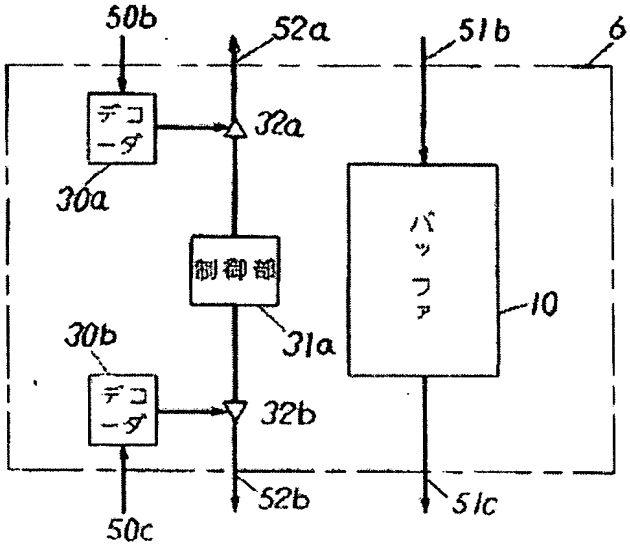


【図１１】

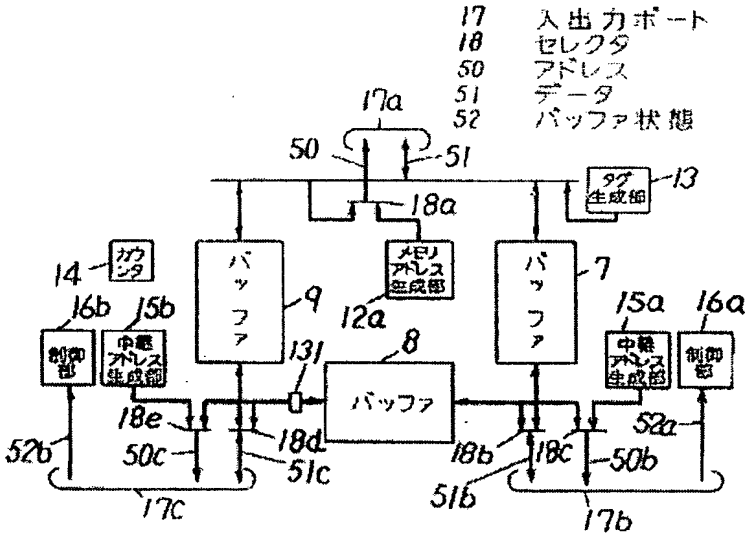
17a.17b17c　入出力ポート



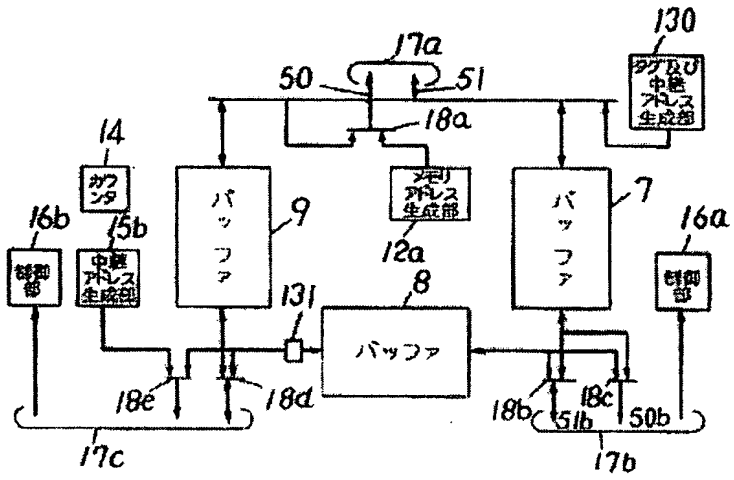
【図１２】


50b
52a
51b
6
デコーダ
32a
30a
制御部
バッファ
10
31a
30b
デコーダ
32b
52b
51c
50c

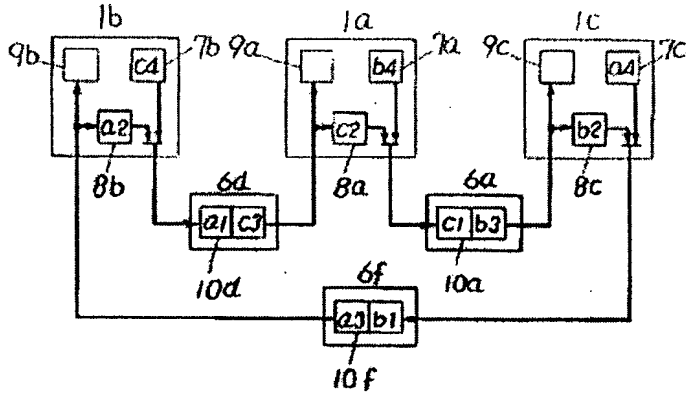
【図１５】


17 入出力ポート
18 セレクタ
50 アドレス
51 データ
52 バッファ状態
17a
50
51
18a
13
14
カウンタ
バッファ
9
メモリアドレス生成部
12a
バッファ
7
15a
16a
16b
15b
制御部
中継アドレス生成部
131
8
バッファ
18e
18d
18b
18c
52a
52b
50c
51c
51b
50b
17c
17b

【図１６】

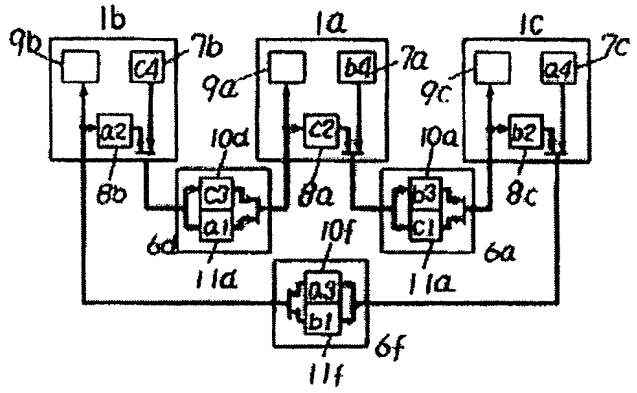

17a
130
50
51
18a
14
16b
15b
9
12a
7
16a
131
8
バッファ
18e
18d
18b
18c
51b
50b
17c
17b

【図１７】

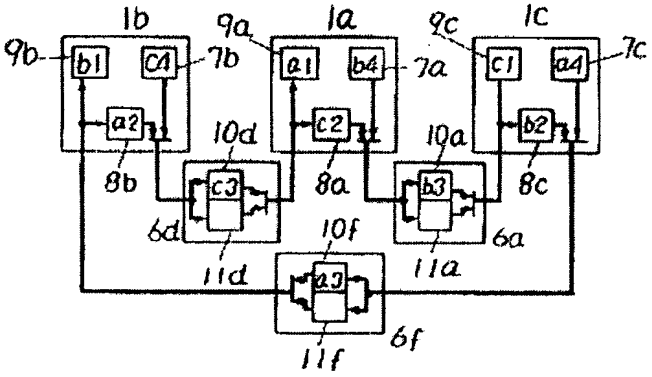

1b
1a
1c
9b
7b
9a
7a
9c
7c
c4
b4
a4
a2
c2
b2
8b
8a
8c
6d
6a
a1
c3
c1
b3
10d
10a
6f
a3
b1
10f

【図18】

(a)


1b
1a
1c
9b
7b
c4
9a
b4
7a
9c
a4
7c
a2
10d
c2
10a
b2
8b
c3
8a
b3
8c
a1
6d
10f
c1
6a
11a
11a
a3
b1
6f
11f

(b)


1b
9a
1a
9c
1c
9b
b1
c4
7b
a1
b4
7a
c1
a4
7c
a2
10d
c2
10a
b2
8b
c3
8a
b3
8c
6d
10f
6a
11d
a3
11a
6f
11f

【図19】